Entropic Control Improvisation for Prediction, Inference, and, Testing

Marcell Vazquez-Chanlatte Sebastian Junges Daniel J. Fremont Sanjit A. Seshia

CAV 20'. SGs preprint.

Entropic Control Improvisation

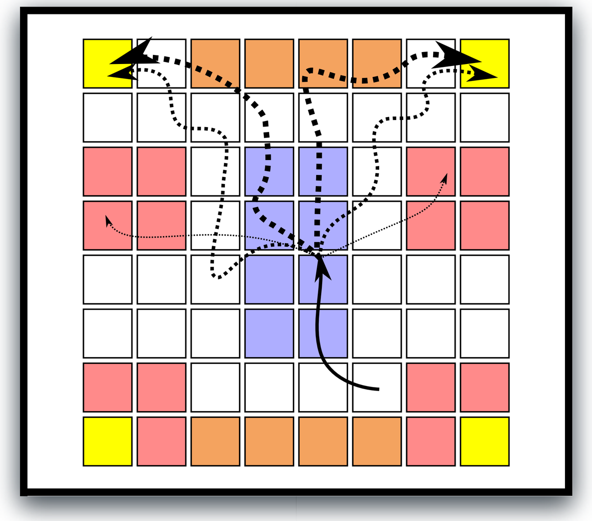
Given a dynamics model and horizon T, find a policy that satisfies:

Hard Constraint:	$\Pr(\xi \in \psi) \geq 1$
Soft Constraint:	$\Pr(\xi \in \varphi) \geq \mathbf{p}$
Causal Entropy Constraint:	$H(\mathcal{A}_{1:T} \mid\mid \mathcal{S}_{1:T}) \geq \mathbf{h}$

1. Causal entropy (**randomness**) constraint ensures minimal bias.
2. Natural trade off between performance **p** and randomness **h**.

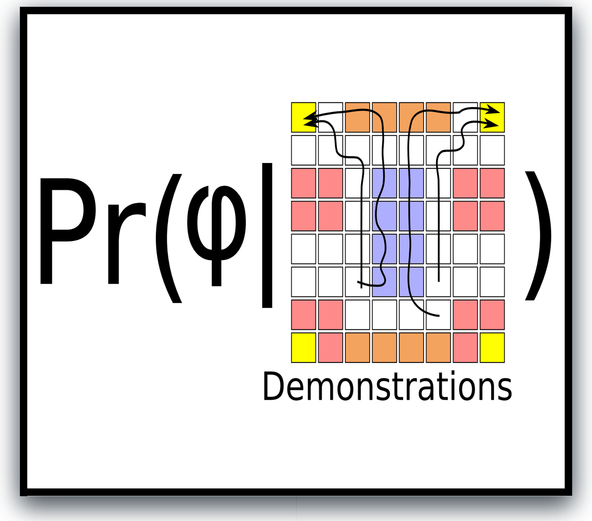
Applications

Prediction



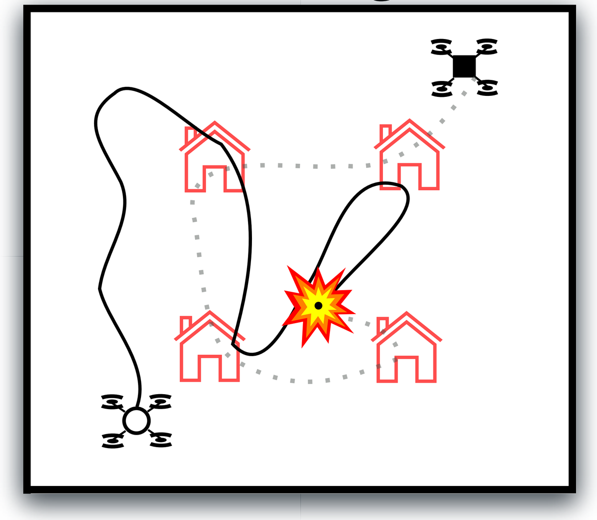
Given φ, predict agents next actions.

Inference



Given demonstrations, predict φ.

Testing



Declaratively specify environment for testing.

In above settings, a biased policy is **undesirable**.

Contributions

1. Algorithm for learning temporal constraints from **unlabeled** demonstrations in Markov Decision Processes (CAV 20').
2. Symbolic approach for representing MDPs and Stochastic Games as **Binary Decision Diagrams** (CAV 20').
3. Improvisation in stochastic games which support arbitrary **combinations** of probabilistic 🎲 and adversarial 👿 uncertainty (In submission).

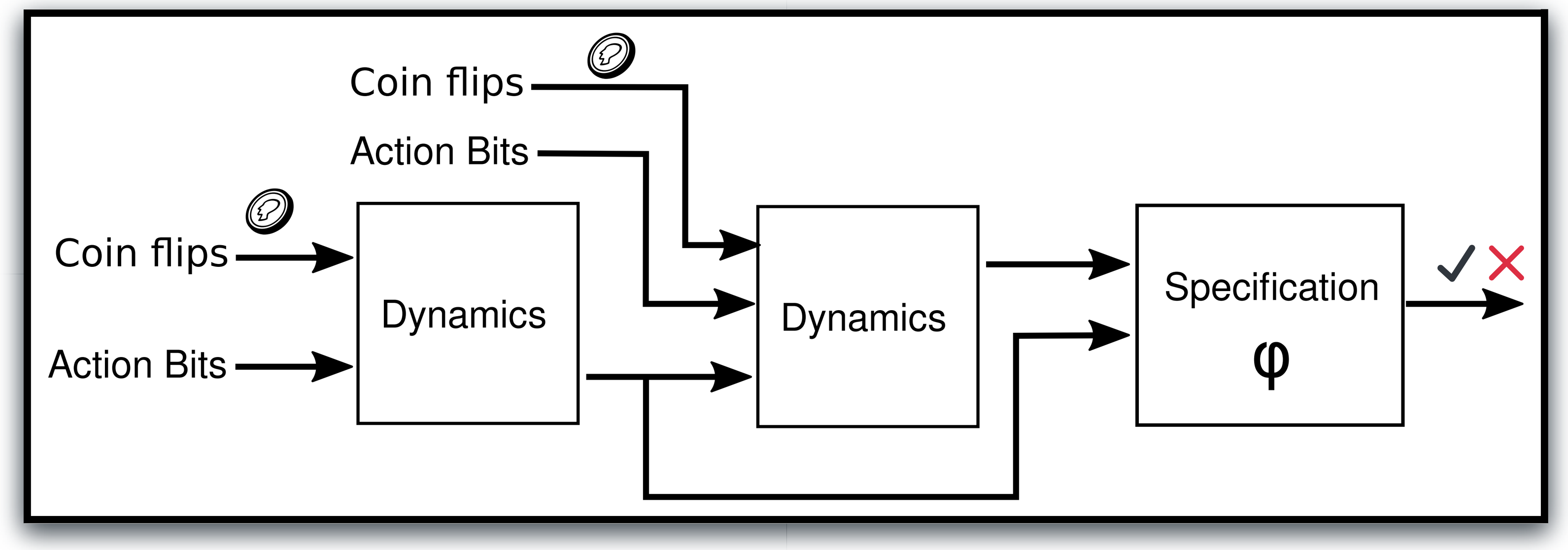
Improv in MDPs (CAV 20')

Key Observation: Can think of soft constraint as binary reward.

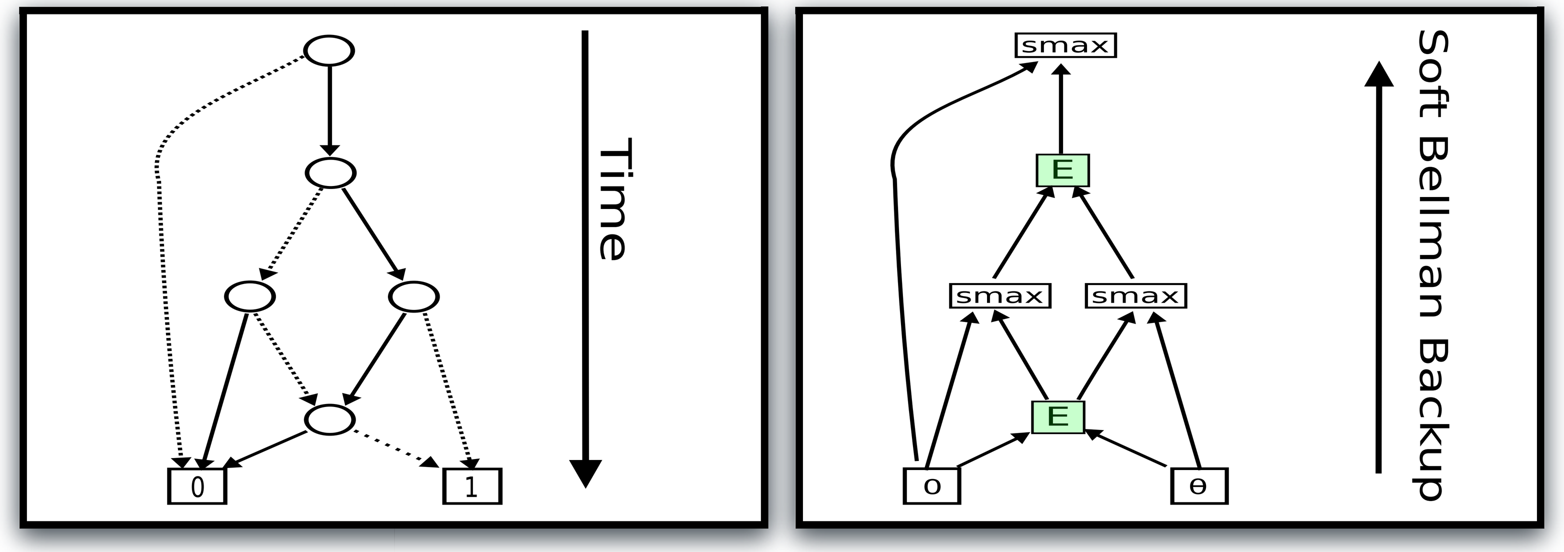
$$r_\lambda(\xi) \triangleq \lambda \cdot 1[\xi \in \varphi]$$

- By adding history to state space, can reduce to Maximum Causal Entropy Inverse Reinforcement Learning.
- **Problem:** Potential combinatorial explosion.
- **Solution:** Encode MDP as a Binary Decision Diagram.

1. Write the **composition** of the dynamics and property as a circuit with access to biased coins.



2. Can represent MDP with a Binary Decision Diagram:



Conservative size bound:

$$O(|\text{horizon}| \cdot |S/\varphi| \cdot |\text{Actions}| \log(|\text{Actions}|))$$

3. We show you can efficiently compute maximum causal entropy policy on compressed MDP.

Application: Used to learn temporal logic constraint from **unlabeled** demonstrations, e.g.,

φ = "Avoid Lava, eventually recharge, and don't recharge while wet."

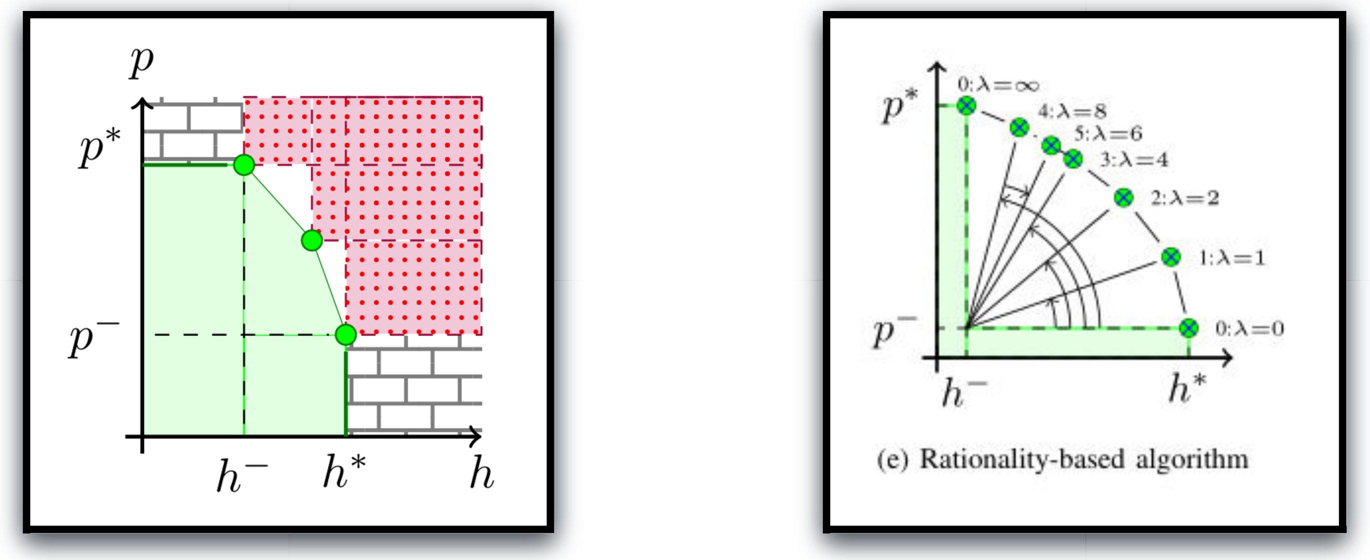
Improv in Stochastic Games (In Submission)

Motivation: Often want to handle combinations of probabilistic 🎲 and adversarial 👿 uncertainty, i.e., Interval MDPs, 2 player MDPs, and model compression.

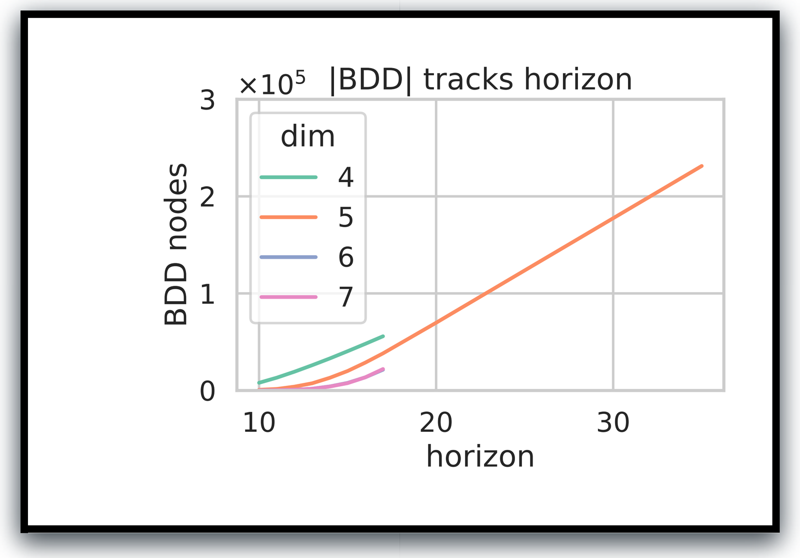
Q: Is efficent improvisation synthesis possible?
A: Yes! By **recursive entropy matching**.

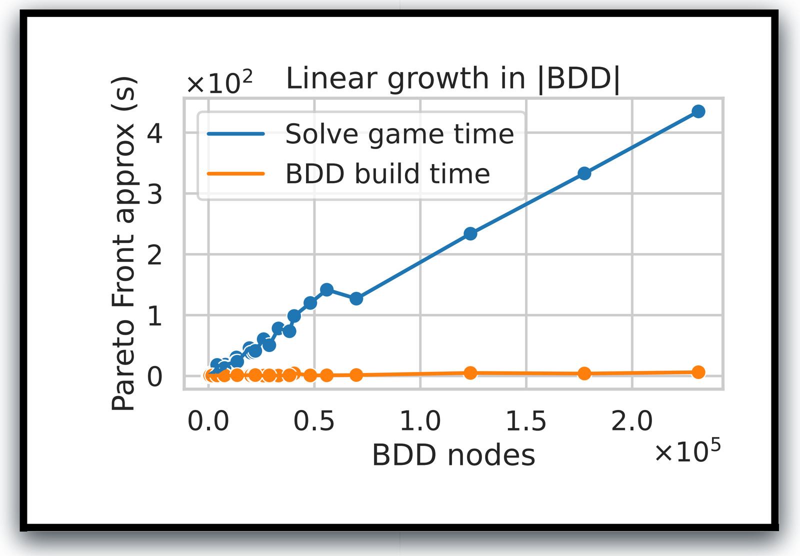
Efficient via dynamic programming from leafs of BDD.

1. Assume min entropy 👿.
2. Run MDP to find optimal **h**.
3. Plan to match **h**.
4. Approximate Pareto Front.



1. Pareto Front allows for re-planning locally.
2. Resulting algorithm is efficient in practice.





Future Work

1. Sampling based algorithms.
2. POMDPs.
3. Subset queries.
4. Dynamic Scenic Constraints.